U.S. Senate Committee

on

Commerce, Science and Transportation

U.S. Senate Committee Testimony, March 28, 2001

Room SR-253

TESTIMONY OF EUGENE P. ERICKSEN

Phone: 215-204-1446 Fax: 215-864-3840
Email: erickgen@astro.temple.edu

Temple University

1.   Introduction

I am a Professor of Sociology and Statistics at Temple University, where I have taught since 1971.   I teach courses on survey design and methods, general statistics, and demography. Last year, I taught a graduate seminar on the United States census, with an emphasis on the nature, causes, and consequences of census error.

I completed my doctoral dissertation at the University of Michigan in 1971. In this dissertation, I developed a method for calculating local population estimates when census data are unavailable. The Census Bureau provided financial support for this work. Since completing the dissertation, I have done substantial research on methods of calculating population estimates, both in census and non-census years. Over the past two decades, I have extended my research into the area of estimating local undercounts.

In October 1980 I advised plaintiffs in a lawsuit filed by the City and State of New York attempting to compel an adjustment of the 1980 Census. Between October 1989 and July 1991, I served as Co-Chair of a Special Advisory Panel appointed by then-Secretary of Commerce Robert Mosbacher to advise him on the possibility of adjusting the 1990 Census. After completing my duties on the Advisory Panel, I again advised plaintiffs in a lawsuit seeking to compel the adjustment of the 1990 Census. For both the 1980 and the 1990 Census, I conducted substantial amounts of research on data relevant to the undercount.

In June 2000, the Census Monitoring Board, Presidential members, hired me as a statistical consultant.   My duties have included reviewing Census Bureau plans and reports regarding the 2000 Census, and analyzing certain data from the census that the Bureau has provided to us. In particular, I have had access to the individual level data records of persons included in the Accuracy and Coverage Evaluation (A.C.E.) samples used to estimate the levels of undercounting and overcounting in the 2000 Census.

2.   Basic Census Concepts

There have been three major controversies of the 2000 Census. The first, settled by a Supreme Court decision, was whether sampling could be used as part of Non-

Response Follow-Up (NRFU). The second, decided by Secretary of Commerce Donald Evans, was whether the results of the census should be adjusted to correct for the differential undercount of minorities for redistricting purposes. The third, still undecided, is whether the adjusted results should be released for public use. I understand that the Census Bureau has already calculated block level adjusted counts for the entire nation.

These controversies are intimately related. To discuss them, it helps to define two terms, "net undercount" and "gross coverage error." The net undercount is the difference between the number of people counted in the census and the bureau's independently calculated estimate of the national population. The net undercount was 1.6 percent in 1990. The bureau estimates that it fell to 1.2 percent, or 3.3 million people, in 2000.

These 3.3 million people are not the total number of people missed from the census. The net undercount is the difference between two quantities, omissions and erroneous inclusions. Omissions are people who should have been counted, but were not. Erroneous inclusions are counts that should not have occurred. They are frequently duplications of the same person counted in the same place. They may also be double counts at separate addresses. For example, college students may be counted not only at the dormitory where they actually live, but also by their parents at home. A family with a second home, perhaps used for weekends and vacations, may get counted at each address. It is conceivable that there could be large, but equal numbers of omissions and erroneous inclusions. Should this occur, the net undercount would be zero, and useless for the evaluation of census error. If the omissions and erroneous inclusions occur at different locations, some local areas would have overcounts and others would have undercounts even though the national net undercount was zero. For example, in 1990, the Census Bureau estimated a 7.76 rate of **undercount** for Black renters living in New York City. In that same census, there was an **overcount** of 0.23 percent among "non-Hispanic White and Other" homeowners also living in New York City[1].

There were several million erroneous inclusions in both the 1990 and the 2000 Censuses. Since these must be subtracted from the number of omissions to derive the net undercount, the actual number of people missed from the census is much greater than the

---

[1] Howard Hogan, "The 1990 Post-Enumeration Survey: Operations and Results," Journal of the American Statistical Association, 88: 1047-1060, Table A.1.

net undercount. The Census Bureau has asserted that 8.4 million persons were omitted from the 1990 Census. The comparable number for the 2000 Census is between 6.4 and 7.6 million persons (see Table 1).

The "gross coverage error" is the sum of omissions and erroneous inclusions. In 1990, the Bureau told us that there were 8.4 million omissions and 4.4 million erroneous inclusions. The net undercount was 4.0 million and gross coverage error was 12.8 million. In 2000, the Bureau indicates that the net undercount was 3.3 million and the gross coverage error was between 9.5 and 11.9 million people.

The national net undercount is not a good indicator of overall census quality. If omissions and erroneous inclusions occurred in the same places, many of them would offset each other. Because the geographic distributions of omissions and erroneous inclusions differ, they offset each other only partially. The gross coverage error is a better indicator of census quality.

Omissions tend to be concentrated among poor, typically minority, populations. They occur at especially high rates among the urban poor. Erroneous inclusions, while present among the poor, frequently occur among the affluent. In 1990, there were net overcounts among Whites living in owner-occupied housing units in the Northeast. In 2000, there were net overcounts among White homeowners in several post-strata located in the Northeast and Midwest.

Census taking is controversial due to the perception that some groups are disadvantaged because of counting errors. The best statistic for evaluating this possibility is the differential undercount. This is the difference in the net undercount between White and minority populations (see Table 2). In 1990, the net undercount for Hispanics was 5.0 percent, for non-Hispanic Whites it was 0.7 percent, so the differential undercount between Hispanics and non-Hispanic Whites was 4.3 percentage points. The net undercount for non-Hispanic Blacks was 4.6 percent and the differential between non-Hispanic Blacks and Whites was 3.9 percentage points.

The results of the 2000 Census show considerable improvement in this statistic. The net undercount for Hispanics was 2.8 percent, for non-Hispanic Blacks it was 2.2 percent, and for non-Hispanic Whites it was 0.7 percent. The differential between Hispanics and non-Hispanic Whites was 2.1 percent and between non-Hispanic Blacks

and Whites it was 1.5 percent. These differentials for the 2000 Census are less than half of the comparable numbers for the 1990 Census. This is real progress, and I congratulate the Census Bureau.

However, from the data I have seen to date, it is clear to me that statistical correction of the census would improve the accuracy. There were a considerable number of errors in this census – at least 9.5 million by the Bureau's own account. The Bureau reduced the numbers of omissions more than the numbers of erroneous enumerations, and therefore reduced the net undercount. Although it reduced the differential undercount as well, important racial differences remain[2], and these could be removed by statistical correction.

3.   The Status of the Controversies

The Supreme Court has settled the question of whether sampling should be used for Non-response Follow-up. We do not need to return to it now. The Secretary of Commerce, following the recommendation of the Census Bureau, has declared the unadjusted count to be the official result of the 2000 Census. We similarly do not need to return to this question. The remaining controversy concerns the release of adjusted block counts.

I believe that these should be released. There are three main reasons for this. One is that while the Census Bureau did reduce all of the net undercount, gross coverage error, and differential undercount rates -- important differentials remain. The American Indian undercount, while less than it was 1990, is at 5 percent. Both the American public and the scientific community need to understand the effects of the undercounts on the census results they rely upon.

The second reason is that there are important ways in which the published census results appear to be incorrect. Ironically, the very improvement in census coverage that the Bureau accomplished creates error. There is confusion between the amount of actual growth and the amount of coverage improvements. Many localities, especially those with large minority populations, may have inflated impressions of the 1990 – 2000 growth

---

[2] These differentials are found whether we rely upon the A.C.E. or demographic analysis.

rates. The best way to fix this problem is to compare adjusted 1990 to adjusted 2000 census estimates. The use of adjusted data will reduce the inconsistencies created by improved coverage in the 2000 Census.

The third reason is that both the American public and the scientific community need to have access to the adjusted counts to make their own evaluations on the nature of census error. There will be some purposes for which statisticians and other data analysts may deem it best to use the unadjusted counts, and other purposes for which they will prefer the adjusted results. The Bureau needs to release the adjusted data, along with their evaluations of these data, to permit these considerations.

We also need to have a better understanding of the geographic distributions of gross error. Study of the gross error will help us to learn where there were important problems of census taking, and where improvements might have taken place in the 2000 compared to the 1990 Census. For example, did the addition of addresses through LUCA reduce omissions? Finally, the scientific community needs to make its own evaluation of the bureau's estimates of the gross error rates.

4.   State and Local Undercounts

To illustrate the kinds of calculations and evaluations scientists need to make, I have calculated undercount estimates for each of the 50 states and the District of Columbia. I have also done this for five large cities, Atlanta, Chicago, New York City, Philadelphia, and Houston, and the remainders of the states in which they are located. I relied upon the A.C.E. data provided by the Census Bureau to the Census Monitoring Board. Because I do not have access to the P.L.94-171 census count data, I cannot match the estimates of these quantities that the Census Bureau would calculate. I should be close, however.

The state estimates vary across a narrow range. All but 10 states have estimates within one-half percentage point of the national average, 1.2 percent, i.e., and they are in the range of 0.7 to 1.7 percent. States with larger shares of minority population tend to have higher rates of undercount, and the opposite occurs in states with smaller minority shares.   Holding the minority shares constant, rates of undercount were higher in the

West and lower in the Midwest. The main result, though, is that between-state variations are moderate.

Turning to the cities, I have calculated rates of undercount for five of them. Although each of the cities includes large minority populations, the rates of undercount are lower than in 1990. Each of these cities has a higher rate of undercount than the remainder of the state in which it is located.

These calculations are possible with the limited amounts of data that the Census Bureau provided to the Census Monitoring Board. To calculate comparable estimates for smaller areas, and to calculate the state and large city estimates with greater certainty, we need to have the adjusted block level data.

5.   Estimates of Population Growth

With the reduction in the net undercount, especially of minority populations, we have difficulty interpreting rates of local population growth. Use of the currently unadjusted data leads to comparison problems. Because the level of undercount, especially in minority areas, was so much greater in 1990 than it is in 2000, use of the unadjusted results of the two censuses leads us to overestimate population growth. The problem is well illustrated by the case of New York City. Since 1990, Census Bureau population estimates have indicated small increases, with a total growth of 1.4 percent occurring between 1990 and 1999. Extrapolating to 2000, the expected population count was 7,452,184, an increase of 1.8 percent since 1990.

 The actual 2000 Census count was 8,008,278, an increase of 9.4 percent since 1990. This amount is nearly 7 percentage points greater than the expectation.   The result is implausible. Local experts believe that New York City has grown, but probably not by this amount. We cannot tell how much of the change is due to improved census coverage, and how much is real growth. In addition, because New York City added several hundred thousand addresses during the LUCA, this operation alone could account for much of the apparent growth. LUCA was not part of the 1990 Census.

The New York City story was repeated consistently across the nation. I have calculated 2000 population projections simply by extrapolating the 1998 – 99 population

change forward to 2000. In Philadelphia, for example, the Bureau estimated a loss of 17,367 people between 1998 and 1999. I simply assumed a comparable loss between 1999 and 2000. The resulting projection of 1,400,234 is 7.7 percentage points below the 2000 Census count of 1,517,550.

I have made comparable calculations for all counties that had at least a 500,000 population in 1990 for which the Census Bureau has published 2000 Census tabulations of P.L.94 – 171 data (see Table 3). There are 66 of these counties and 60 of them have counts that are higher than the population projections than I calculated. Only 6 have counts that are below the projections. This skewed pattern of difference reflects the fact that the 2000 Census counts are more complete than those of the 1990 Census.

I have divided the counties into four categories, depending on the percentage minority in the 2000 Census. Among those counties where this percentage was less than 20, the average difference between the projected estimates and actual counts was moderate, 1.72 percent. Fourteen of the 16 population estimates were too low.

Turning to the second category, where the percentage minority was between 20 and 29.9 percent, there were 22 counties, and 19 of them had estimates that were too low. The average difference between the population projection and the actual count was 2.35 percent.

The third category included 16 counties where the percentage minority was between 30 and 49.9 percent. Fifteen of them had projections that were too low. The average difference between the population projection and the actual count was 3.78 percent.

The fourth and final category included 12 counties where the percentage minority was greater than 50 percent. All of these counties had projections that were too low, and the average difference between the projection and the actual count was 6.66 percent. These counties were spread across the country, with 5 of them located in the Northeast, 6 of them in the South, and 1 in the West.

Overall, we see that the discrepancies between projections and counts are greater where the percentage minority is greater. To understand these confusing patterns of population growth and coverage improvement, we need to have both the adjusted and unadjusted block counts. With these in hand, we can create our own combinations of

local jurisdictions to study the extent to which the apparent growth is real or the product of changed and improved methods of census taking.

6. Summary

The 2000 Census is now complete, and its surprising results require analysis. The Census Bureau must complete its evaluation of the demographic estimate, and assess its consistency with the A.C.E. results. Demographers and statisticians across the country recognize this as a crucial project, and they would like to have their own looks at the data. In addition to understanding the predictors of undercount in the 2000 Census, the changing patterns of undercount between the 1990 and 2000 Censuses, we need to figure out why the demographic and A.C.E. results are so discrepant.

We also need to understand the patterns of gross and net error. Although I have not been able to carry out intensive research on the subject, much of the improvement in census coverage appears to be due to improvements in the census address list. The Local Update to Census Addresses program was a big contributor to this. If this were true, we would expect a substantial reduction in the number of "whole household omissions" in the 2000 Census relative to 1990. If the address list is improved, then the number of entire households missed should go down.

On the other hand, many omissions occur in households where others are counted. These "within household omissions" occur when people filling out census forms misunderstand the instructions, and such omissions are especially prevalent among persons distantly or unrelated to the heads of household. Because such errors are generally unrelated to the completeness of address lists, or even to the proficiency of advertising and other outreach programs, we would expect the numbers of such omissions to be similar to those observed in 1990.

Finally, we need good estimates of the numbers of omissions and erroneous inclusions. The Census Bureau has estimated that there were over 20 million non-matches, but only 6.4 to 7.6 true omissions in the A.C.E.. The Bureau also estimates that only 3.1 to 4.3 million of the 12 million apparent erroneous inclusions are real. Many of the apparent errors are due to missing data and other similar problems, and the numbers

of apparent errors overstate the problems. However, the Census Bureau has not yet shared with us the logic behind these estimates.

There is a great deal of research that needs to be done to understand the 2000 Census. Some of it must necessarily be carried out at the Census Bureau, but the bureau must make public all data relevant to this understanding so that researchers and academicians can draw their own conclusions.